Prometheus: Unsupervised Discovery of Phase Transitions Through Physics-Informed Variational Autoencoders

Brandon Yee[1*], **Wilson Collins**[1], **Caden Wang**[2], **Mihir Tekal**[1]

[1]Yee Collins Research Group, Weston, CT 06883, USA
[2]College of Arts and Sciences, New York University, New York, NY 10003, USA
[*]Corresponding author; b.yee@ycrg-labs.org

Abstract

Applying AI to scientific problems requires addressing domain-specific challenges including physical constraint integration, interpretability requirements, and theoretical validation. We present Prometheus, a physics-informed variational autoencoder that demonstrates successful AI-science collaboration through unsupervised discovery of phase transitions in the 2D Ising model. Our approach achieves 0.85 correlation with theoretical order parameters and 99.7% critical temperature accuracy by incorporating symmetry constraints, progressive training, and physics-informed loss functions. This work provides a template for interdisciplinary research, offering concrete strategies for integrating domain knowledge with AI techniques and establishing validation protocols for scientific discovery.

Code — https://github.com/YCRG-Labs/prometheus

1 AI-Science Collaboration: Challenges and Solutions

Applying AI to scientific problems faces unique challenges that differ from typical machine learning applications (Carrasquilla and Melko 2017). In condensed matter physics, phase transition discovery traditionally requires prior knowledge of order parameters (Landau and Lifshitz 1980), presenting specific AI challenges:

Physical Constraints: Physics problems require adherence to symmetries and conservation laws that standard neural networks may violate through spurious correlations.

Interpretability: Scientific applications demand interpretable results validated against theoretical predictions, not just predictive accuracy.

Limited Supervision: Many phenomena lack labeled datasets, requiring unsupervised approaches for meaningful pattern extraction.

This work addresses the workshop's objectives by: (1) identifying specific challenges in applying deep learning to phase transitions, (2) developing physics-informed VAE adaptations, and (3) demonstrating successful AI-physics collaboration methodologies.

2 Physics-Informed Methodology

Problem Formulation and Domain Analysis

The 2D Ising model provides an ideal testbed with known theoretical properties (Onsager 1944). The system exhibits a continuous phase transition at $T_c = 2/\ln(1+\sqrt{2}) \approx 2.269$ with Hamiltonian:

$$H = -J\sum_{\langle i,j\rangle} s_i s_j - h\sum_i s_i \tag{1}$$

Our AI-science collaboration challenge is to discover these relationships automatically while respecting physical principles.

Architecture Design Principles

Our physics-informed VAE integrates domain knowledge through systematic design choices:

Symmetry-Preserving Architecture: We implement convolutional layers that respect the $\mathbb{Z}_2$ spin-flip symmetry and lattice translation invariance. The encoder uses symmetric 3×3 kernels with periodic padding, while data augmentation includes rotations (90°, 180°, 270°) and reflections that preserve Ising physics.

Latent Space Design: The 8-dimensional latent space captures both local spin correlations and global order parameters through hierarchical representation learning.

Progressive Training Strategy: Our curriculum learning approach mirrors the physical phase diagram:

1. **High-Temperature Phase** ($T > T_c$)**:** Initial training on disordered configurations ($T \in [2.5, 3.5]$) establishes basic representational capacity for random spin arrangements.
2. **Critical Region** ($T \approx T_c$)**:** Gradual introduction of near-critical configurations ($T \in [2.1, 2.4]$) enhances sensitivity to correlation length divergence and critical fluctuations.
3. **Low-Temperature Phase** ($T < T_c$)**:** Final training on ordered states ($T \in [1.5, 2.1]$) captures domain formation and spontaneous symmetry breaking.

Physics-Informed Loss Function Design

The total loss function incorporates multiple physics-based terms addressing specific domain challenges:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \beta\mathcal{L}_{KL} + \lambda_1\mathcal{L}_{symmetry} + \lambda_2\mathcal{L}_{energy} + \lambda_3\mathcal{L}_{correlation} \tag{2}$$

where: - $\mathcal{L}_{symmetry} = \|\mu(x) - \mu(-x)\|_2^2$ enforces $\mathbb{Z}_2$ symmetry in latent representations - $\mathcal{L}_{energy}$ penalizes reconstructions with unphysical energy distributions - $\mathcal{L}_{correlation}$ encourages proper spatial correlation structure - Hyperparameters $\beta = 1.0$, $\lambda_1 = 0.1$, $\lambda_2 = 0.05$, $\lambda_3 = 0.02$ balance objectives

Automated Order Parameter Discovery Protocol

We develop a systematic protocol for identifying physical order parameters from learned representations:

1. **Correlation Analysis:** Compute Pearson correlations between each latent dimension and known physical quantities (magnetization, energy, specific heat).
2. **Temperature Dependence:** Analyze how latent dimensions vary with temperature, identifying those exhibiting critical behavior near T_c.
3. **Symmetry Analysis:** Verify that discovered order parameters transform correctly under physical symmetries.
4. **Finite-Size Scaling:** Validate that discovered quantities exhibit proper scaling behavior across different lattice sizes.

3 Comprehensive Results and Validation

Experimental Setup and Dataset

We generate 50,000 Monte Carlo configurations using the Metropolis-Hastings algorithm with careful equilibration protocols. Temperature sampling covers $T \in [1.5, 3.5]$ with 50 temperature points, ensuring dense coverage around the critical region. Lattice sizes $L \in \{16, 32, 64\}$ enable finite-size scaling analysis.

Performance Comparison and Statistical Analysis

Our physics-informed VAE significantly outperforms standard unsupervised methods:

Table 1: Method comparison results

Method	Order Parameter Correlation	Critical Temp Error (%)
PCA	0.45 ± 0.08	12.3
Standard VAE	0.62 ± 0.06	8.1
β-VAE	0.68 ± 0.07	6.8
Physics-Informed VAE	**0.85 ± 0.04**	**0.27**

Table 2: Ablation study results

Configuration	Order Parameter Correlation	Critical Temp Error (%)
Full Framework	0.85 ± 0.04	0.27
w/o Physics Constraints	0.72 ± 0.08	3.2
w/o Progressive Training	0.78 ± 0.06	1.8
w/o Symmetry Augmentation	0.81 ± 0.05	0.9

Statistical significance testing using bootstrap resampling (n=1000) confirms substantial improvements: physics-informed VAE vs. standard VAE (p ¡ 0.001, Cohen's d = 2.8), demonstrating both statistical and practical significance.

Detailed Physics Validation

Order Parameter Discovery: Our automated protocol successfully identifies latent dimension 3 as correlating most strongly with magnetization (r = 0.85, p ¡ 0.001). The discovered order parameter exhibits proper temperature dependence: $m(T) \propto (T_c - T)^\beta$ with critical exponent $\beta = 0.1240.008$, matching theoretical prediction $\beta = 1/8$.

Critical Temperature Detection: Using susceptibility peak analysis on the discovered order parameter, we detect $T_c = 2.2630.006$, representing 0.27% error from the theoretical value. The method maintains accuracy across lattice sizes with proper finite-size corrections.

Finite-Size Scaling Analysis: The discovered order parameter exhibits correct finite-size scaling: $m_L(T_c) \propto L^{-\beta/\nu}$ with $\beta/\nu = 0.1250.012$, consistent with theoretical prediction 0.125.

Interpretability and Scientific Insight

Latent Space Structure: Analysis reveals hierarchical organization: dimensions 1-2 capture local spin correlations, dimensions 3-4 encode global magnetization, and dimensions 5-8 represent higher-order correlations and critical fluctuations.

Phase Boundary Identification: The learned representations naturally separate ordered and disordered phases in latent space, with clear geometric structure requiring no manual threshold selection.

4 Lessons for AI-Science Collaboration

Successful Integration Strategies

Our interdisciplinary experience reveals key principles for effective AI-science collaboration:

Domain Knowledge Integration: Incorporating physical principles as architectural constraints and loss terms significantly improves both accuracy and interpretability. The 89% improvement over PCA demonstrates that domain-informed approaches substantially outperform generic methods.

Validation Against Theory: Rigorous comparison with theoretical predictions builds confidence in AI discoveries. Our critical exponent validation ($\beta = 0.1240.008$ vs. theoretical 0.125) exemplifies how AI results can be scientifically validated.

Interpretable Representations: Designing latent spaces aligned with physical concepts enables scientific insight beyond predictive accuracy. Our hierarchical latent organization provides interpretable structure matching physical understanding.

Progressive Complexity: Curriculum learning strategies mirroring scientific understanding accelerate training and improve generalization.

Challenges and Practical Solutions

Challenge: Balancing Physics Constraints and Model Flexibility *Solution:* Adaptive weighting schemes that adjust physics constraint strength during training, preventing over-regularization while maintaining physical validity.

Challenge: Validating Discoveries in Unknown Systems *Solution:* Multi-level validation protocols combining theoretical consistency checks, symmetry analysis, and cross-validation with known limiting cases.

Challenge: Communicating AI Results to Domain Scientists *Solution:* Develop visualization tools and interpretability metrics that translate AI discoveries into domain-familiar concepts and terminology.

Transferable Methodological Framework

This work establishes a general framework applicable across scientific domains:

Step 1: Domain Analysis - Identify key symmetries, conservation laws, and theoretical constraints specific to the scientific problem.

Step 2: Architecture Design - Incorporate domain knowledge through network architecture, data augmentation, and loss function design.

Step 3: Progressive Training - Develop curriculum learning strategies that mirror scientific understanding and problem complexity.

Step 4: Automated Discovery - Implement systematic protocols for identifying scientifically meaningful patterns in learned representations.

Step 5: Rigorous Validation - Establish comprehensive validation against theoretical predictions, experimental data, and known limiting cases.

Cross-Domain Applications and Case Studies

Materials Science: Our symmetry-preserving architecture principles apply to crystal structure analysis, where space group symmetries constrain atomic arrangements. The progressive training approach can be adapted to learn structure-property relationships from simple to complex materials.

Climate Science: The automated discovery protocol translates to identifying climate patterns and tipping points. Physical constraints include energy conservation and atmospheric dynamics, while validation involves comparison with climate models and observational data.

Astronomy: Stellar classification benefits from physics-informed approaches incorporating stellar evolution theory and spectroscopic constraints. The hierarchical latent organization can capture multi-scale phenomena from stellar surfaces to galactic structures.

Molecular Biology: Protein folding prediction can incorporate physical constraints from thermodynamics and structural biology. Progressive training from simple to complex fold families mirrors evolutionary complexity.

Community Building Insights

Interdisciplinary Communication: Success requires developing shared vocabulary and conceptual frameworks through regular joint seminars and collaborative workshops.

Educational Integration: Training programs should combine AI methodology with domain science fundamentals through 6-month collaborative projects.

Open Science Practices: Reproducible research practices, including open-source code and comprehensive documentation, accelerate community adoption and enable collaborative validation of results.

5 Community Building and Future Directions

Workshop Recommendations for AI-Science Integration

Based on our interdisciplinary experience and workshop objectives, we provide actionable recommendations:

1. Establish Domain-Informed AI Development Protocols Create standardized frameworks for incorporating scientific knowledge into AI architectures. Our physics-informed VAE demonstrates 89% performance improvement, suggesting systematic domain integration yields substantial benefits across scientific applications.

2. Develop Rigorous Validation Standards Implement multi-level validation protocols combining theoretical consistency, experimental verification, and cross-domain validation. Our critical exponent validation exemplifies how AI discoveries can meet scientific rigor standards.

3. Prioritize Scientific Interpretability Design AI systems that provide scientific insight, not just predictive accuracy. Our hierarchical latent organization demonstrates how interpretable representations enable scientific understanding beyond correlation analysis.

4. Foster Interdisciplinary Education and Training Develop structured programs combining AI methodology with domain science. Recommended curriculum includes: (a) 3-month AI fundamentals for scientists, (b) 3-month domain science immersion for AI researchers, (c) 6-month collaborative project implementation.

5. Build Collaborative Infrastructure Create platforms facilitating ongoing AI-science partnerships, including shared computational resources, standardized datasets, and collaborative validation protocols.

Immediate Next Steps for the Community

Short-term (6 months): - Organize monthly AI-science collaboration seminars - Establish shared repositories for physics-informed AI architectures - Create validation benchmark datasets for scientific AI applications

Medium-term (1-2 years): - Develop interdisciplinary graduate programs and postdoctoral fellowships - Launch collaborative funding initiatives supporting AI-science partnerships - Establish peer review standards for interdisciplinary AI-science publications

Long-term (3-5 years): - Create international consortiums for large-scale AI-science projects - Develop standardized certification programs for scientific AI practitioners - Establish dedicated journals for AI-driven scientific discovery

Promising Research Frontiers

Quantum Many-Body Systems: Extension of our physics-informed approach to quantum phase transitions, incorporating quantum entanglement constraints and many-body localization phenomena.

Multi-Scale Modeling: Integration across length and time scales, from molecular dynamics to continuum mechanics using hierarchical latent organization.

Real-Time Scientific Discovery: Online learning systems for real-time experimental data analysis and hypothesis generation. Applications include adaptive experimental design and autonomous scientific discovery.

Causal Scientific Discovery: Moving beyond correlation to identify causal relationships in complex systems. Integration of causal inference methods with physics-informed AI promises breakthrough capabilities in understanding scientific causation.

Automated Theory Generation: AI systems that generate testable scientific hypotheses and theoretical frameworks. Our automated order parameter discovery provides a foundation for more general theory generation capabilities.

Broader Impact on Scientific Practice

Democratization of Scientific Discovery: Physics-informed AI tools can enable researchers without extensive theoretical backgrounds to make meaningful scientific contributions, broadening participation in scientific research.

Acceleration of Scientific Progress: Automated discovery protocols can dramatically reduce the time from data collection to scientific insight, potentially accelerating scientific progress by orders of magnitude.

Enhanced Reproducibility: Standardized AI-science frameworks improve reproducibility by providing systematic methodologies and validation protocols that can be consistently applied across research groups.

Cross-Disciplinary Fertilization: Success in one scientific domain can rapidly transfer to others through shared methodological frameworks, enabling cross-pollination of ideas and techniques.

Addressing Potential Challenges

Quality Control: Establish peer review standards that properly evaluate both AI methodology and scientific validity. Recommend interdisciplinary review panels with expertise in both domains.

Ethical Considerations: Develop guidelines for responsible AI use in scientific discovery, including proper attribution, validation requirements, and transparency standards.

Resource Allocation: Create funding mechanisms that support long-term interdisciplinary collaborations rather than short-term projects, recognizing that meaningful AI-science integration requires sustained effort.

Cultural Integration: Address cultural differences between AI and scientific communities through structured interaction programs and shared success metrics that value both technological innovation and scientific insight.

6 Conclusions and Workshop Impact

This work demonstrates that successful AI-science collaboration transcends simple application of existing AI tools to scientific problems. Instead, it requires systematic integration of domain knowledge, rigorous theoretical validation, and unwavering commitment to interpretable results that advance scientific understanding.

Our physics-informed VAE exemplifies these principles through unsupervised discovery of phase transitions, achieving 0.85 correlation with theoretical order parameters while maintaining full scientific interpretability. The 89% improvement over traditional methods (Cohen's d = 2.8) demonstrates both statistical significance and practical impact, establishing a new standard for AI-driven scientific discovery.

Workshop Goal Achievement: This work directly advances the workshop's three primary objectives: (1) *Challenge Identification* through systematic analysis of AI-science integration difficulties, (2) *Tool Development* via physics-informed VAE innovations and automated discovery protocols, and (3) *Community Building* through transferable methodologies and actionable recommendations for interdisciplinary collaboration.

Methodological Contributions: We establish a comprehensive framework for AI-science integration comprising domain analysis, physics-informed architecture design, progressive training strategies, automated discovery protocols, and rigorous validation methodologies. This framework provides a template for future collaborations across diverse scientific domains.

Community Impact: Our interdisciplinary approach offers concrete strategies for bridging AI and scientific communities, including educational recommendations, infrastructure development, and collaborative protocols that can be immediately implemented by workshop participants and the broader community.

The success of Prometheus validates the workshop's central thesis: AI can meaningfully advance scientific discovery when developed with careful attention to domain requirements, theoretical foundations, and community needs. By providing both technical innovations and practical guidance for collaboration, this work establishes a foundation for the next generation of AI-driven scientific breakthroughs.

Acknowledgments

We thank the condensed matter physics and machine learning communities for valuable discussions. This work demonstrates the power of open-source software and reproducible research practices in enabling interdisciplinary collaboration.

References

Carrasquilla, J.; and Melko, R. G. 2017. Machine learning phases of matter. *Nat Phys*, 13(5): 431–434.

Landau, L. D.; and Lifshitz, E. M. 1980. *Statistical Physics*. Oxford: Pergamon Press, 3rd edition.

Onsager, L. 1944. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys Rev*, 65(3-4): 117–149.